Epistemic Game Theory

Lecture 5

ESSLLI'12, Opole

Eric Pacuit Olivier Roy

TiLPS, Tilburg University MCMP, LMU Munich
`ai.stanford.edu/~epacuit`
`http://olivier.amonbofis.net`

August 10, 2012

Plan for the week

1. **Monday** Basic Concepts.
2. **Tuesday** Epistemics.
3. **Wednesday** Fundamentals of Epistemic Game Theory.
4. **Thursday** Tree, Puzzles and Paradoxes.
5. **Friday** More Puzzles, Extensions and New Directions.
    - Admissibility continued.
    - The Brandenburger-Kiesler Paradox.
    - Nash Equilibrium?
    - Concluding remarks.

LPS: $(\mu_0, \mu_1, \ldots, \mu_{n-1})$ (each μ_i is a probability measure with disjoint supports)

LPS: $(\mu_0, \mu_1, \ldots, \mu_{n-1})$ (each μ_i is a probability measure with disjoint supports)

(s_i, t_i) is **rational** provided (i) s_i lexicographically maximizes i's expected payoff under the LPS associated with t_i, **and** (ii) the LPS associated with t_i has full support.

LPS: $(\mu_0, \mu_1, \ldots, \mu_{n-1})$ (each μ_i is a probability measure with disjoint supports)

(s_i, t_i) is **rational** provided (i) s_i lexicographically maximizes i's expected payoff under the LPS associated with t_i, **and** (ii) the LPS associated with t_i has full support.

A player **assumes** E provided she considers E infinitely more likely than *not*-E.

LPS: $(\mu_0, \mu_1, \ldots, \mu_{n-1})$ (each μ_i is a probability measure with disjoint supports)

(s_i, t_i) is **rational** provided (i) s_i lexicographically maximizes i's expected payoff under the LPS associated with t_i, **and** (ii) the LPS associated with t_i has full support.

A player **assumes** E provided she considers E infinitely more likely than *not*-E.

The key notion is **rationality and common assumption of rationality** (RCAR).

But, there's more...

"Under admissibility, Ann considers everything possible. But this is only a decision-theoretic statement. Ann is in a game, so we imagine she asks herself: "What about Bob? What does he consider possible?" If Ann truly considers everything possible, then it seems she should, in particular, allow for the possibility that Bob does not! Alternatively put, it seems that a full analysis of the admissibility requirement should include the idea that other players do not conform to the requirement." (pg. 313)

A. Brandenburger, A. Friedenberg, H. J. Keisler. *Admissibility in Games*. Econometrica (2008).

Irrationality

			2	
		L	C	R
1	T	4,0	4,1	0,1
	M	0,0	0,1	4,1
	D	3,0	2,1	2,1

Irrationality

			2	
		L	C	R
1	T	4,0	4,1	0,1
	M	0,0	0,1	4,1
	D	3,0	2,1	2,1

- The IA set

Irrationality

		2		
		L	C	R
1	T	4,0	4,1	0,1
	M	0,0	0,1	4,1
	D	3,0	2,1	2,1

- All (L, b_i) are irrational, (C, b_i), (R, b_i) are rational if b_i has full support, irrational otherwise
- D is optimal then either $\mu(C) = \mu(R) = \frac{1}{2}$ or μ assigns positive probability to both L and R.

Irrationality

			2	
		L	C	R
1	T	4,0	4,1	0,1
	M	0,0	0,1	4,1
	D	3,0	2,1	2,1

- Fix a rational (D, a) where a assumes that Bob is rational. $(a \mapsto (\mu_0, \ldots, \mu_{n-1}))$
- Let μ_i be the first measure assigning nonzero probability to $\{L\} \times T_B$ ($i \neq 0$ since a assumes Bob is rational).

Irrationality

			2	
		L	C	R
1	T	4,0	4,1	0,1
	M	0,0	0,1	4,1
	D	3,0	2,1	2,1

- Let μ_i be the first measure assigning nonzero probability to $\{L\} \times T_B$ ($i \neq 0$).
- for each μ_k with $k < i$: (i) μ_k assigns probability $\frac{1}{2}$ to $\{C\} \times T_B$ and $\frac{1}{2}$ to $\{R\} \times T_B$; and (ii) U, M, D are each optimal under μ_k.

Irrationality

			2	
		L	C	R
1	T	4,0	4,1	0,1
	M	0,0	0,1	4,1
	D	3,0	2,1	2,1

- for each μ_k with $k < i$: (i) μ_k assigns probability $\frac{1}{2}$ to $\{C\} \times T_B$ and $\frac{1}{2}$ to $\{R\} \times T_B$; and (ii) T, M, D are each optimal under μ_k.
- D must be optimal under μ_i and so μ_i assigns positive probability to both $\{L\} \times T_B$ and $\{R\} \times T_B$.

Irrationality

		2		
		L	C	R
1	T	4,0	4,1	0,1
	M	0,0	0,1	4,1
	D	3,0	2,1	2,1

- D must be optimal under μ_i and so μ_i assigns positive probability to both $\{L\} \times T_B$ and $\{R\} \times T_B$.
- Rational strategy-type pairs are each infinitely more likely that irrational strategy-type pairs. Since, each point in $\{L\} \times T_B$ is irrational, μ_i must assign positive probability to irrational pairs in $\{R\} \times T_B$.

Irrationality

		2		
		L	C	R
1	T	4,0	4,1	0,1
	M	0,0	0,1	4,1
	D	3,0	2,1	2,1

- μ_i must assign positive probability to irrational pairs in $\{R\} \times T_B$.
- *This can only happen if there are types of Bob that do not consider everything possible.*

The Brandenburger-Keisler Paradox

		2		
		l	c	r
1	t	4,4	1,1	0,0
	m	1,1	5,5	0,0
	d	0,1	0,1	6,0

	b
l	1
c	0
r	0

	a
t	1
m	0
d	0

Brandenburger-Kiesler Paradox

		2		
		l	c	r
1	t	4,4	1,1	0,0
	m	1,1	5,5	0,0
	d	0,1	0,1	6,0

	b
l	1
c	0
r	0

	a
t	1
m	0
d	0

- The projection of $RCBR$ is $\{(t, l)\}$

			2	
		l	c	r
1	t	4,4	1,1	0,0
	m	1,1	5,5	0,0
	d	0,1	0,1	6,0

	b
l	1
c	0
r	0

	a
t	1
m	0
d	0

- The projection of *RCBR* is $\{(t, l)\}$
- This is not the entire *ISDS* set

Brandenburger-Kiesler Paradox

2

		l	c	r
1	t	4,4	1,1	0,0
	m	1,1	5,5	0,0
	d	0,1	0,1	6,0

	b
l	1
c	0
r	0

	a
t	1
m	0
d	0

- The projection of $RCBR$ is $\{(t, l)\}$
- This is not the entire $ISDS$ set
- "Game independent" conditions and *rich* type structures

A Question

- For any given set S of external states we can use a Bayesian model or a type space on S to provide consistent representations of the players' beliefs.

A Question

- For any given set S of external states we can use a Bayesian model or a type space on S to provide consistent representations of the players' beliefs.
- Every state in a belief model or type space induces an infinite hierarchy of beliefs, but *not all consistent and coherent infinite hierarchies are in any finite model*. It is not obvious that even in an infinite model that all such hierarchies of beliefs can be represented.

A Question

- For any given set S of external states we can use a Bayesian model or a type space on S to provide consistent representations of the players' beliefs.
- Every state in a belief model or type space induces an infinite hierarchy of beliefs, but *not all consistent and coherent infinite hierarchies are in any finite model*. It is not obvious that even in an infinite model that all such hierarchies of beliefs can be represented.
- Which type space is the "correct" one to work with?

Some Literature

A. Brandenburger and E. Dekel. *Hierarchies of Beliefs and Common Knowledge*. Journal of Economic Theory (1993).

Some Literature

A. Brandenburger and E. Dekel. *Hierarchies of Beliefs and Common Knowledge*. Journal of Economic Theory (1993).

A. Heifetz and D. Samet. *Knoweldge Spaces with Arbitrarily High Rank*. Games and Economic Behavior (1998).

Some Literature

A. Brandenburger and E. Dekel. *Hierarchies of Beliefs and Common Knowledge*. Journal of Economic Theory (1993).

A. Heifetz and D. Samet. *Knoweldge Spaces with Arbitrarily High Rank*. Games and Economic Behavior (1998).

L. Moss and I. Viglizzo. *Harsanyi type spaces and final coalgebras constructed from satisfied theories*. EN in Theoretical Computer Science (2004).

Some Literature

A. Brandenburger and E. Dekel. *Hierarchies of Beliefs and Common Knowledge*. Journal of Economic Theory (1993).

A. Heifetz and D. Samet. *Knoweldge Spaces with Arbitrarily High Rank*. Games and Economic Behavior (1998).

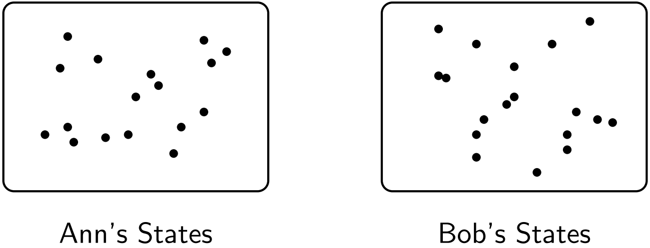
L. Moss and I. Viglizzo. *Harsanyi type spaces and final coalgebras constructed from satisfied theories*. EN in Theoretical Computer Science (2004).

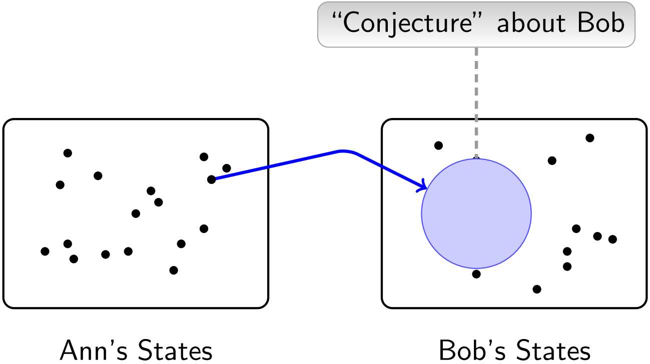
A. Friendenberg. *When do type structures contain all hierarchies of beliefs?*. working paper (2007).

The General Question

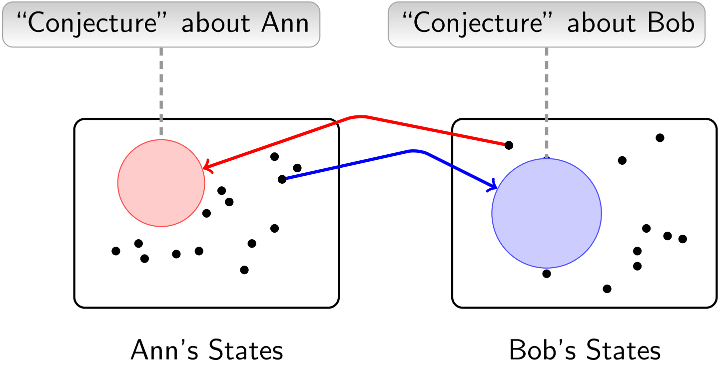
Does there exist a space of "*all possible*" beliefs?

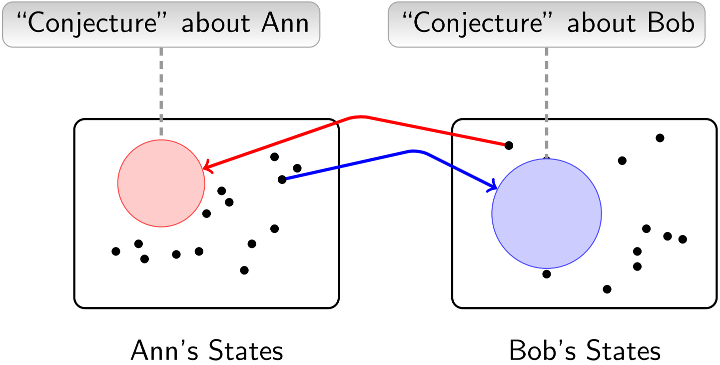
Brandenburger-Kiesler Paradox



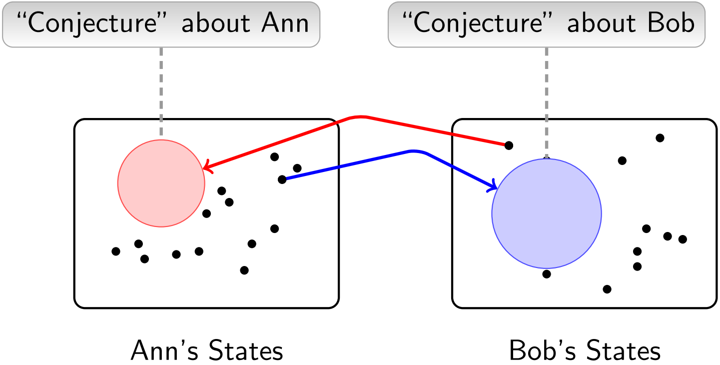

"Conjecture" about Bob
Ann's States
Bob's States

Brandenburger-Kiesler Paradox





Is there a space where every *possible* conjecture is considered by *some* type?



Is there a space where every *possible* conjecture is considered by *some* type? It depends...

A Paradox

Ann believes that Bob assumes* that
Ann believes that Bob's assumption is wrong.

Does Ann believe that Bob's assumption is wrong?

$*$ An **assumption** (or strongest belief) is a belief that implies all other beliefs.

A. Brandenburger and H. J. Keisler. *An Impossibility Theorem on Beliefs in Games*. *Studia Logica* (2006).

A Paradox

Ann believes that Bob assumes* that
Ann believes that Bob's assumption is wrong.

Does Ann believe that Bob's assumption is wrong? Yes.

A Paradox

Ann believes that Bob assumes* that
Ann believes that Bob's assumption is wrong.

Does Ann believe that Bob's assumption is wrong? Yes.

Then according to Ann, Bob's assumption is right.

A Paradox

Ann believes that Bob assumes* that
Ann believes that Bob's assumption is wrong.

Does Ann believe that Bob's assumption is wrong? Yes.

Then according to Ann, Bob's assumption is right.

But then, Ann does not believe Bob's assumption is wrong.

A Paradox

Ann believes that Bob assumes* that
Ann believes that Bob's assumption is wrong.

Does Ann believe that Bob's assumption is wrong? Yes.

Then according to Ann, Bob's assumption is right.

But then, Ann does not believe Bob's assumption is wrong.

So, the answer must be no.

A Paradox

Ann believes that Bob assumes* that
Ann believes that Bob's assumption is wrong.

Does Ann believe that Bob's assumption is wrong? No.

A Paradox

Ann believes that Bob assumes* that
Ann believes that Bob's assumption is wrong.

Does Ann believe that Bob's assumption is wrong? No.

Then Ann does not believe that Bob's assumption is wrong.

A Paradox

Ann believes that Bob assumes* that
Ann believes that Bob's assumption is wrong.

Does Ann believe that Bob's assumption is wrong? No.

Then Ann does not believe that Bob's assumption is wrong.

Then, in Ann's view, Bob's assumption is wrong.

A Paradox

Ann believes that Bob assumes* that
Ann believes that Bob's assumption is wrong.

Does Ann believe that Bob's assumption is wrong? No.

Then Ann does not believe that Bob's assumption is wrong.

Then, in Ann's view, Bob's assumption is wrong.

So, the answer must be yes.

S. Abramsky and J. Zvesper. *From Lawvere to Brandenburger-Keisler: interactive forms of diagonalization and self-reference*. Proceedings of LOFT 2010.

EP. *Understanding the Brandenburger Keisler Pardox*. Studia Logica (2007).

Impossibility Results

Language: the (formal) language used by the players to formulate conjectures about their opponents.

Impossibility Results

Language: the (formal) language used by the players to formulate conjectures about their opponents.

Completeness: A model is **complete for a language** if every (consistent) statement in a player's language about an opponent is *considered* by some type.

Brandenburger-Kiesler Paradox

Qualitative Type Spaces: $\langle T_a, T_b, \lambda_a, \lambda_b \rangle$

$\lambda_a : T_a \to \wp(T_b)$
$\lambda_b : T_b \to \wp(T_a)$

Brandenburger-Kiesler Paradox

Qualitative Type Spaces: $\langle T_a, T_b, \lambda_a, \lambda_b \rangle$

$\lambda_a : T_a \to \wp(T_b)$
$\lambda_b : T_b \to \wp(T_a)$

x **believes** a set $Y \subseteq T_b$ if $\{y \mid y \in \lambda_a(x)\} \subseteq Y$

x **assumes** a set $Y \subseteq T_b$ if $\{y \mid y \in \lambda_a(x)\} = Y$

Impossibility Results

Impossibility 1 There is no complete interactive belief structure for the *powerset language*.

Proof. Cantor: there is no onto map from X to the nonempty subsets of X.

Impossibility Results

Impossibility 1 There is no complete interactive belief structure for the *powerset language*.

Proof. Cantor: there is no onto map from X to the nonempty subsets of X.

Impossibility 2 (Brandenburger and Keisler) There is no complete interactive belief structure for *first-order logic*.

Suppose that $\mathcal{C}_A \subseteq \wp(T_A)$ is a set of *conjectures* about Ann and $\mathcal{C}_B \subseteq \wp(T_B)$ a set of conjectures about Bob states.

Assume For all $X \in \mathcal{C}_A$ there is a $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$: "in state x_0, Ann has consistent beliefs"
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$: "in state x_0, Ann believes that Bob assumes X"

Lemma. Under the above assumption, for each $X \in \mathcal{C}_A$ there is an x_0 such that

$$x_0 \in X \text{ iff there is a } y \in T_B \text{ such that } y \in \lambda_A(x_0) \text{ and } x_0 \in \lambda_B(y)$$

Claim. $x_0 \in X$ iff $\exists y \in T_B$, $y \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y)$

For all $X \in \mathcal{C}_A$ there is a $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$

Suppose that $X \in \mathcal{C}_A$. Then there is an $x_0 \in T_A$ satisfying 1 and 2.

Claim. $x_0 \in X$ iff $\exists y \in T_B$, $y \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y)$

For all $X \in \mathcal{C}_A$ there is a $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$

Suppose that $X \in \mathcal{C}_A$. Then there is an $x_0 \in T_A$ satisfying 1 and 2.

Suppose that $x_0 \in X$.

Claim. $x_0 \in X$ iff $\exists y \in T_B$, $y \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y)$

For all $X \in \mathcal{C}_A$ there is a $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$

Suppose that $X \in \mathcal{C}_A$. Then there is an $x_0 \in T_A$ satisfying 1 and 2.

Suppose that $x_0 \in X$. By 1., $\lambda_A(x_0) \neq \emptyset$ so there is a $y_0 \in T_B$ such that $y_0 \in \lambda_A(x_0)$.

Claim. $x_0 \in X$ iff $\exists y \in T_B$, $y \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y)$

For all $X \in \mathcal{C}_A$ there is a $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$

Suppose that $X \in \mathcal{C}_A$. Then there is an $x_0 \in T_A$ satisfying 1 and 2.

Suppose that $x_0 \in X$. By 1., $\lambda_A(x_0) \neq \emptyset$ so there is a $y_0 \in T_B$ such that $y_0 \in \lambda_A(x_0)$. We show that $x_0 \in \lambda_B(y_0)$.

Claim. $x_0 \in X$ iff $\exists y \in T_B$, $y \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y)$

For all $X \in \mathcal{C}_A$ there is a $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$

Suppose that $X \in \mathcal{C}_A$. Then there is an $x_0 \in T_A$ satisfying 1 and 2.

Suppose that $x_0 \in X$. By 1., $\lambda_A(x_0) \neq \emptyset$ so there is a $y_0 \in T_B$ such that $y_0 \in \lambda_A(x_0)$. We show that $x_0 \in \lambda_B(y_0)$. By 2., we have $y_0 \in \lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$. Hence, $x_0 \in X = \lambda_B(y_0)$.

Claim. $x_0 \in X$ iff $\exists y \in T_B$, $y \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y)$

For all $X \in \mathcal{C}_A$ there is a $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$

Suppose that $X \in \mathcal{C}_A$. Then there is an $x_0 \in T_A$ satisfying 1 and 2.

Suppose that $x_0 \in X$. By 1., $\lambda_A(x_0) \neq \emptyset$ so there is a $y_0 \in T_B$ such that $y_0 \in \lambda_A(x_0)$. We show that $x_0 \in \lambda_B(y_0)$. By 2., we have $y_0 \in \lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$. Hence, $x_0 \in X = \lambda_B(y_0)$.

Suppose that there is a $y_0 \in T_B$ such that $y_0 \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y_0)$.

Brandenburger-Kiesler Paradox

Claim. $x_0 \in X$ iff $\exists y \in T_B$, $y \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y)$

For all $X \in \mathcal{C}_A$ there is a $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$

Suppose that $X \in \mathcal{C}_A$. Then there is an $x_0 \in T_A$ satisfying 1 and 2.

Suppose that $x_0 \in X$. By 1., $\lambda_A(x_0) \neq \emptyset$ so there is a $y_0 \in T_B$ such that $y_0 \in \lambda_A(x_0)$. We show that $x_0 \in \lambda_B(y_0)$. By 2., we have $y_0 \in \lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$. Hence, $x_0 \in X = \lambda_B(y_0)$.

Suppose that there is a $y_0 \in T_B$ such that $y_0 \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y_0)$. By 2., $y_0 \in \lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$.

Claim. $x_0 \in X$ iff $\exists y \in T_B$, $y \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y)$

For all $X \in \mathcal{C}_A$ there is a $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$

Suppose that $X \in \mathcal{C}_A$. Then there is an $x_0 \in T_A$ satisfying 1 and 2.

Suppose that $x_0 \in X$. By 1., $\lambda_A(x_0) \neq \emptyset$ so there is a $y_0 \in T_B$ such that $y_0 \in \lambda_A(x_0)$. We show that $x_0 \in \lambda_B(y_0)$. By 2., we have $y_0 \in \lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$. Hence, $x_0 \in X = \lambda_B(y_0)$.

Suppose that there is a $y_0 \in T_B$ such that $y_0 \in \lambda_A(x_0)$ and $x_0 \in \lambda_B(y_0)$. By 2., $y_0 \in \lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$. Hence, $x_0 \in \lambda_B(y_0) = X$.

Consider a first-order language $\mathcal{L}$ containing binary relational symbols $R_A(x, y)$ and $R_B(x, y)$ defining λ_A and λ_B, respectively.

Brandenburger-Kiesler Paradox

Consider a first-order language $\mathcal{L}$ containing binary relational symbols $R_A(x, y)$ and $R_B(x, y)$ defining λ_A and λ_B, respectively.

$\mathcal{L}$ is interpreted over qualitative type structures where the interpretation of R_A is $\{(t, s) \mid t \in T_A, s \in T_B, \text{ and } s \in \lambda_A(t)\}$.

Consider a first-order language $\mathcal{L}$ containing binary relational symbols $R_A(x, y)$ and $R_B(x, y)$ defining λ_A and λ_B, respectively.

$\mathcal{L}$ is interpreted over qualitative type structures where the interpretation of R_A is $\{(t, s) \mid t \in T_A, s \in T_B, \text{ and } s \in \lambda_A(t)\}$.

Consider the formula φ in $\mathcal{L}$:

$$\varphi(x) := \exists y(R_A(x, y) \wedge R_B(y, x))$$

Consider a first-order language $\mathcal{L}$ containing binary relational symbols $R_A(x, y)$ and $R_B(x, y)$ defining λ_A and λ_B, respectively.

$\mathcal{L}$ is interpreted over qualitative type structures where the interpretation of R_A is $\{(t, s) \mid t \in T_A, s \in T_B, \text{ and } s \in \lambda_A(t)\}$.

Consider the formula φ in $\mathcal{L}$:

$$\varphi(x) := \exists y(R_A(x, y) \wedge R_B(y, x))$$

$\neg\varphi(x) := \forall y(R_A(x, y) \rightarrow \neg R_B(y, x))$: "Ann believes that Bob's assumption is *wrong*."

Proof of the Theorem

Suppose that $X \in \mathcal{C}_A$ is defined by the formula $\neg\varphi(x) := \neg\exists y(R_A(x, y) \wedge R_B(y, x))$.

Proof of the Theorem

Suppose that $X \in \mathcal{C}_A$ is defined by the formula $\neg\varphi(x) := \neg\exists y(R_A(x, y) \wedge R_B(y, x))$.

There is an $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$: Ann's beliefs at x_0 are consistent.
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$: At x_0, Ann believes that Bob assumes $X = \{x \mid \neg\varphi(x)\}$ (i.e., Ann believes that Bob assumes that Ann believes that Bob's assumption is wrong.)

Proof of the Theorem

Suppose that $X \in \mathcal{C}_A$ is defined by the formula $\neg\varphi(x) := \neg\exists y(R_A(x,y) \wedge R_B(y,x))$.

There is an $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$: Ann's beliefs at x_0 are consistent.
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$: At x_0, Ann believes that Bob assumes $X = \{x \mid \neg\varphi(x)\}$ (i.e., Ann believes that Bob assumes that Ann believes that Bob's assumption is wrong.)

$\neg\varphi(x_0)$ is true iff (def. of X) $x_0 \in X$

Proof of the Theorem

Suppose that $X \in \mathcal{C}_A$ is defined by the formula $\neg\varphi(x) := \neg\exists y(R_A(x,y) \wedge R_B(y,x))$.

There is an $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$: Ann's beliefs at x_0 are consistent.
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$: At x_0, Ann believes that Bob assumes $X = \{x \mid \neg\varphi(x)\}$ (i.e., Ann believes that Bob assumes that Ann believes that Bob's assumption is wrong.)

$$\begin{array}{lll} \neg\varphi(x_0) \text{ is true} & \text{iff (def. of } X\text{)} & x_0 \in X \\ & \text{iff (Lemma)} & \text{there is a } y \in T_B \text{ with } y \in \lambda_A(x_0) \\ & & \text{and } x_0 \in \lambda_B(y) \end{array}$$

Proof of the Theorem

Suppose that $X \in \mathcal{C}_A$ is defined by the formula
$\neg\varphi(x) := \neg\exists y(R_A(x,y) \wedge R_B(y,x))$.

There is an $x_0 \in T_A$ such that

1. $\lambda_A(x_0) \neq \emptyset$: Ann's beliefs at x_0 are consistent.
2. $\lambda_A(x_0) \subseteq \{y \mid \lambda_B(y) = X\}$: At x_0, Ann believes that Bob assumes $X = \{x \mid \neg\varphi(x)\}$ (i.e., Ann believes that Bob assumes that Ann believes that Bob's assumption is wrong.)

$$\begin{array}{lll}
\neg\varphi(x_0) \text{ is true} & \text{iff (def. of } X\text{)} & x_0 \in X \\
 & \text{iff (Lemma)} & \text{there is a } y \in T_B \text{ with } y \in \lambda_A(x_0) \\
 & & \text{and } x_0 \in \lambda_B(y) \\
 & \text{iff (def. of } \varphi(x)\text{)} & \varphi(x_0) \text{ is true.}
\end{array}$$

- *RCBR* and iterated strict dominance
- *CKRat* and backwards induction
- *RCAR* and iterated weak dominance

Nash Equilibrium

	A	B
a	1, 1	0, 0
b	0, 0	1, 1

- The profiles **aA** and **bB** are two pure-strategy Nash equilibria of that game.

Definition

A strategy profile σ is a *Nash equilibrium* iff for all i and all $s_i' \neq \sigma_i$:

$$u_i(\sigma) \geq u_i(s_i, \sigma_{-i})$$

More Specific Expectations

	A	B
a	1, 1	0, 0
b	0, 0	1, 1

More Specific Expectations

	A	B
a	1, 1	0, 0
b	0, 0	1, 1

- If Ann believes that Bob plays **A**, the only rational choice for her is **a**.

More Specific Expectations

	A	B
a	1, 1	0, 0
b	0, 0	1, 1

- If Ann believes that Bob plays **A**, the only rational choice for her is **a**.
- The same hold for Bob.

More Specific Expectations

	A	B
a	1, 1	0, 0
b	0, 0	1, 1

- If Ann believes that Bob plays **A**, the only rational choice for her is **a**.
- The same hold for Bob.
- If, furthermore, these beliefs are *true*, then **aA** is played.

Knowledge of Strategies and Nash Equilibrium

	A	B
a	1, 1	0, 0
b	0, 0	1, 1

- If Ann and Bob are rational and have correct beliefs about each others' strategy choices, then **aA** is played.

Knowledge of Strategies and Nash Equilibrium

	A	B
a	1, 1	0, 0
b	0, 0	1, 1

- If Ann and Bob are rational and have correct beliefs about each others' strategy choices, then **aA** is played.
- For any two-players strategic game and model for that game, if at state w both players are rational and know the other's strategy choice, then $\sigma(w)$ is a Nash equilibrium.

R. Aumann and A. Brandenburger, "Epistemic Conditions for Nash Equilibrium". *Econometrica*. 1995.

Hard Knowledge of Strategies and Nash Equilibrium

Theorem

(Aumann and Brandenburger, 1995) For any two-players strategic game and model for that game, if at state w both players are rational and know other's strategy choice, then $\sigma(w)$ is a Nash equilibrium.

- Remarks:

Hard Knowledge of Strategies and Nash Equilibrium

Theorem

(Aumann and Brandenburger, 1995) For any two-players strategic game and model for that game, if at state w both players are rational and know other's strategy choice, then $\sigma(w)$ is a Nash equilibrium.

- Remarks:
  - Close to the intuitive explanation: Best response *given* the choices of others, or no regret.

Hard Knowledge of Strategies and Nash Equilibrium

Theorem

(Aumann and Brandenburger, 1995) For any two-players strategic game and model for that game, if at state w both players are rational and know other's strategy choice, then $\sigma(w)$ is a Nash equilibrium.

- Remarks:
  - Close to the intuitive explanation: Best response *given* the choices of others, or no regret.
  - No higher-order information needed... for 2 players (more on this in a moment).

Hard Knowledge of Strategies and Nash Equilibrium

Theorem

(Aumann and Brandenburger, 1995) For any two-players strategic game and model for that game, if at state w both players are rational and know other's strategy choice, then $\sigma(w)$ is a Nash equilibrium.

- Remarks:
  - Close to the intuitive explanation: Best response *given* the choices of others, or no regret.
  - No higher-order information needed... for 2 players (more on this in a moment).
  - Hard knowledge, or even correct beliefs, about actions taken? Does Nash equilibrium undermine strategic uncertainty?

Nash equilibrium, the general case

(Aumann and Brandenburger, 1995) In an n-player game, suppose that the players have a common prior, that their payoff functions and their rationality are mutually known, and that their conjectures are commonly known. Then for each player j, all the other players i agree on the same conjecture σ_j about j, and the resulting profile $(\sigma_1, .., \sigma_n)$ of mixed actions is a Nash equilibrium.

- Remarks:

Nash equilibrium, the general case

(Aumann and Brandenburger, 1995) In an n-player game, suppose that the players have a common prior, that their payoff functions and their rationality are mutually known, and that their conjectures are commonly known. Then for each player j, all the other players i agree on the same conjecture σ_j about j, and the resulting profile $(\sigma_1, .., \sigma_n)$ of mixed actions is a Nash equilibrium.

- Remarks:
  - Higher-order information after all: common knowledge of conjectures.

Nash equilibrium, the general case

(Aumann and Brandenburger, 1995) In an n-player game, suppose that the players have a common prior, that their payoff functions and their rationality are mutually known, and that their conjectures are commonly known. Then for each player j, all the other players i agree on the same conjecture σ_j about j, and the resulting profile $(\sigma_1, .., \sigma_n)$ of mixed actions is a Nash equilibrium.

- Remarks:
  - Higher-order information after all: common knowledge of conjectures.
  - The result is "tight". Fails if we drop any of the conditions.

Nash equilibrium, the general case

(Aumann and Brandenburger, 1995) In an n-player game, suppose that the players have a common prior, that their payoff functions and their rationality are mutually known, and that their conjectures are commonly known. Then for each player j, all the other players i agree on the same conjecture σ_j about j, and the resulting profile $(\sigma_1, .., \sigma_n)$ of mixed actions is a Nash equilibrium.

- Remarks:
  - Higher-order information after all: common knowledge of conjectures.
  - The result is "tight". Fails if we drop any of the conditions.
  - Epistemic Interpretation of mixed strategies.

Nash equilibrium, the general case

(Aumann and Brandenburger, 1995) In an n-player game, suppose that the players have a common prior, that their payoff functions and their rationality are mutually known, and that their conjectures are commonly known. Then for each player j, all the other players i agree on the same conjecture σ_j about j, and the resulting profile $(\sigma_1, .., \sigma_n)$ of mixed actions is a Nash equilibrium.

- Remarks:
  - Higher-order information after all: common knowledge of conjectures.
  - The result is "tight". Fails if we drop any of the conditions.
  - Epistemic Interpretation of mixed strategies.
  - If the payoffs are common knowledge, then rationality is also common knowledge (Ben Polak, Econometrica, 1999).

Nash equilibrium, the general case

(Aumann and Brandenburger, 1995) In an n-player game, suppose that the players have a common prior, that their payoff functions and their rationality are mutually known, and that their conjectures are commonly known. Then for each player j, all the other players i agree on the same conjecture σ_j about j, and the resulting profile $(\sigma_1, .., \sigma_n)$ of mixed actions is a Nash equilibrium.

- Remarks:
  - Higher-order information after all: common knowledge of conjectures.
  - The result is “tight”. Fails if we drop any of the conditions.
  - Epistemic Interpretation of mixed strategies.
  - If the payoffs are common knowledge, then rationality is also common knowledge (Ben Polak, Econometrica, 1999).
  - But still, CKR does not imply Nash Equilibrium.

Some Concluding Remarks

Some Concluding Remarks

Common Knowledge of Rationality

Common Knowledge of Rationality

- Variety of individual attitudes: Beliefs, conditional beliefs, safe/robust beliefs, strong beliefs, lexical probability systems...

Common Knowledge of Rationality

- Variety of individual attitudes: Beliefs, conditional beliefs, safe/robust beliefs, strong beliefs, lexical probability systems...
- Different modes of collective attitudes: mutual beliefs, finite levels, distributed knowledge...

Common Knowledge of Rationality

- Variety of individual attitudes: Beliefs, conditional beliefs, safe/robust beliefs, strong beliefs, lexical probability systems...
- Different modes of collective attitudes: mutual beliefs, finite levels, distributed knowledge...
- Different choice rules: admissibility, minmax, minmax Regret, more abstract notions...

Common Knowledge of Rationality

- Variety of individual attitudes: Beliefs, conditional beliefs, safe/robust beliefs, strong beliefs, lexical probability systems...
- Different modes of collective attitudes: mutual beliefs, finite levels, distributed knowledge...
- Different choice rules: admissibility, minmax, minmax Regret, more abstract notions...

In which direction to go?

- Towards normatively plausible theories.
- Towards descriptively adequate theories.

These need not always to be different directions, or at least independent from one another...

Some Concluding Remarks

The point of view of this model is not normative; it is not meant to advise the players what to do. The players do whatever they do; their strategies are taken as given.

Some Concluding Remarks

The point of view of this model is not normative; it is not meant to advise the players what to do. The players do whatever they do; their strategies are taken as given. Neither is it meant as a description of what human beings actually do in interactive situations.

Some Concluding Remarks

The point of view of this model is not normative; it is not meant to advise the players what to do. The players do whatever they do; their strategies are taken as given. Neither is it meant as a description of what human beings actually do in interactive situations. The most appropriate term is perhaps "analytic"; it asks, what are the implications *of rationality in interactive situations? Where does it lead?*

The point of view of this model is not normative; it is not meant to advise the players what to do. The players do whatever they do; their strategies are taken as given. Neither is it meant as a description of what human beings actually do in interactive situations. The most appropriate term is perhaps "analytic"; it asks, what are the implications *of rationality in interactive situations? Where does it lead? This question may be as important as, or even more important than, more direct "tests" of the relevance of the rationality hypothesis.*

R. Aumann. *Irrationality in Game Theory*. 1992.

Some Concluding Remarks

Thank you for listening!